Let the AI assisted medicine remain human
Developing a care centered model of ethical decision making in AI based healthcare

Francesca Morpurgo [1], Carmela Occhipinti [2]

[1] *CyberEthicsLab Srls, Rome, Italy*

Abstract

How to reconcile an algorithmic, Artificial Intelligence (AI) based approach to healthcare with a truly humanistic attitude to medical research and practice? How to ensure that the outcomes of AI based medical systems comply with "ethical" principles and that can be used in an ethically sound way?

This paper presents here the first version of a model for ethical decision making in AI-based healthcare developed in the context of the EU funded research project MES-CoBraD, that aims at developing an AI-based decision support system for the diagnosis and treatment of complex brain diseases. The model, called ETHAI (Ethical AI) is meant to guide both the clinicians making ethical decisions while using the project AI-based expert system in their daily practice, and the AI-based expert system developers to make an ethics-compliant output. The ETHAI model is based on a set of ethics requirements grounded in the real medical practice and with a solid technical feasibility, to be validated in the course of the project.

Keywords

Principlism, care ethics, bioethics, artificial intelligence, AI, ethical decision models, healthcare, research ethics

1. Introduction

Artificial Intelligence (AI) is becoming more and more present in healthcare, especially in some fields, where it has a fundamental part in the diagnostic process. While it certainly is something that introduces extremely positive developments (e.g. early identification of diseases), it doesn't come without a price (just to name one, the risk of introducing biases or misdiagnoses at scale).

Especially when the AI intervention implies a certain amount of data elaboration and decision-making, basing medical decisions and diagnoses on the output of the AI system, it has strong and not yet fully solved ethical implications, that deepen when fragile people or particularly complex diseases are involved.

This opens a huge debate on how to ensure that ethics constraints and requirements are respected when using a form of AI, both in research and in real healthcare scenarios. Of course, from the Oviedo Convention on and passing through the classical four principles of biomedical ethics (beneficence, non-maleficence, autonomy and justice), there are quite strict rules and requirements to ensure that any biomedical research or healthcare activity is conducted by agreed-upon ethics principles.

However**,** the crucial point is that when AI is applied into a specific field, new issues to be tackled might emerge, e.g. the potential presence of biases in datasets on which the algorithm is trained**,** the question of the accountability**,** the explicability and fairness of the outputs**,** the human autonomy that could be endangered from

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29-31, 2023, Pisa, Italy
EMAIL: f.morpurgo@cyberethicslab.com (A. 1); c.occhipinti@cyberethicslab.com (A. 2)

an AI-approach**,** and the issue of data privacy and protection always on the background**.**[2]

Thus, even if all the stated requirements are respected, there might be reasons that bring to a misalignment between what should come as an ethics-respecting outcome of such an AI based healthcare system and what actually happens. For this reason, it is extremely important that decision-making processes are designed and carried out in a way that guarantees the respect of ethics requirements and principles. That is even more true when we consider the interaction aspect: healthcare professionals will have to use the AI system, interact with it and its outputs, deal with patients' doubts and fears about the use of AI and, in some cases, decide how much weight to give to a diagnosis coming from the algorithm.

Two different but co-related goals should be pursued to tackle these concerns:

- ensuring that the process itself without AI is ethically compliant;
- ensuring that the process with AI continues to be ethically compliant, with respect to both the outputs and the interaction point of view.

A model for AI-based ethical decision making should ensure both: having the first as its basis, it should ensure that the introduction of AI into the diagnostic process continues to deliver ethically compliant outputs.

2. The need for an ethical AI model

The concept itself of what "ethical" means is at stake here: since introducing AI in the process may imply deep changes in the way patients are cared for as well as in the doctor/patient relationship, a different approach in the ethics of AI is needed, an approach that takes into account fundamental human needs that can't be definitely forgotten or neglected.

Should something that while improves certain aspects of care represents a worsening of others be considered "ethical"? Is the risk of an overly reductionist approach - that relegates the patient only to an amount of data to be processed by an algorithm, resulting in a sort of datafication of the subject - likely to occur? Would it be possible to reconcile a "care" approach to healthcare with an engineering-based one?

The debate about these questions is broad and quite lively. Maio G. (2018) for instance detects the difference between traditional deductive ethics and care ethics in the accent that in the latter is put on a form of "implicit knowledge" which implies the role of the relationships, the correct perception of the situation, the reliance on experiential, situational and relationship knowledge going beyond what is defined and accepted in a formal-logical approach. Barnes et al (2015) also qualifies as central to this approach the understanding of the relational nature of human beings.

Of course, in this approach a purely deterministic stance would result in the loss of some fundamental aspects of healthcare, therefore - while certainly improving certain areas, for instance delivering faster or more accurate diagnoses - it would cause a generalized worsening of the patient's experience.

These are highly relevant questions for the research that is being undertaken inside MES-CoBraD (Multidisciplinary Expert System for the Assessment & Management of Complex Brain Disorders - https://www.mes-cobrad.eu/), an EU funded project that is working on a new protocol for diagnosing and caring for complex brain diseases (epilepsy, dementia, insomnia, Alzheimer).

In MES-CoBraD, 14 partners from 10 different countries and research centers are working together to exploit the potential of data and AI to develop a common innovative protocol for the diagnosis and care of complex brain diseases.

The project implies many ethically sensitive areas: recruiting participants in the clinical study - that due to the nature of the field are often fragile, if not mentally impaired, people -, extracting their data and making them available for all the medical partners of the consortium, using an AI-based system to analyse patients' data and help clinicians in taking a decision regarding

[2] For a thorough overview of the issues deriving from this, see Morley et al (2019), Murphy, K et al (2021) and Karimian, G et al (2022)

their diagnosis and therapy, finding a way for both patients and clinicians to interact with the platform without being spoiled or disregarded or ignored.

In order to assess and cope with these ethics issues, the project defined its own model that, thanks to the research nature of the project, can be effectively tested on the field. This will allow the project to validate it experimentally, stepping from the purely theoretical side to the practical one, with the advantage of a continuous confrontation with medical researchers and practitioners, as well as - at a certain step of the project - with patients and/or stakeholders.

3. Description of the research currently being done

While the work on developing the model has not been completed yet, its first version has been already applied to assess the first results of the project, mainly regarding patients' recruitment, data storage and analysis, as well as the design of the expert system that, using training datasets and a form of supervised machine learning will - at the end of the project - support medical professionals in their work.

As part of the assessment, the consortium partners (a composite mixture of healthcare professionals, researchers and engineers) have been asked to respond to some questions mainly about i) how they are dealing with some key questions like participants recruitment and personal data protection, ii) how they perceive the relationship between doctor and patient, iii) how they feel about the healthcare professionals' accountability in the light of the presence of the expert systems outputs.

From this preliminary work, some interesting open points to reflect upon emerged:

- Explicability: how to make access to the "reasoning" of the system and to understand it.
- Trustability, strictly intertwined with the robustness and with the potential presence of biases.
- Fairness and Autonomy: who decides about the validity of a certain output and on what basis, and who is accountable for it.
- Ethics of Care: how can we ensure that i) our systems and the protocols connected to them remain "human", and ii) we respect not only major basic principles but also the need of the patient to be considered and cared for.

Questions such as how to ensure the accessibility and understandability of the system's reasoning, the position to assume about the clinicians' autonomy and accountability (how can a clinician be deemed responsible for basing his/her decision on a diagnosis he/she cannot fully understand? And if the patient is to be considered as the ultimate subject of autonomy, how can his/her decisions be considered autonomous if they are based on opaque data and "reasoning"?), the risk of depersonalisation of care, the validity and trustworthiness of the system's results, the effects that such a system may have when used at scale, and the possibility of biases and of misuses (e.g. the chance to use it for social scoring purposes) emerge strongly from this line of research and call for accurately meditated answers.

These open questions introduce into the landscape a strong ethical uncertainty and call for the need of a way of supporting both healthcare professionals and developers in the process of building and interacting with an expert system dedicated to healthcare, pointing out the need of an ethics-compliant AI model.

As a consequence, this brings again to the original question of what is considered to be ethical and of what an "ethical" model should do.

4. The ETHAI model

First of all, it is important to reflect on:

- what a "model" is
- what it is for
- what form such a model should assume,
- how it could represent an advantage for healthcare professionals and developers.

In the MES-CoBraD approach, the developed ETHAI (Ethical AI) model has two functions: on the one hand it has to support clinicians who will use the expert system in their daily practice, offering a guide when they face an ethical dilemma or problem (generated or not by the "intrusion" of AI in the diagnostic process); on the other hand, it has to represent a guide for the

developers working on the expert system, so that they know how to build a platform that with a good amount of probability delivers ethically sound outputs. There could also be a third function that in the MES-COBRAD project does not apply and represents an advancement of the current model: acting as a guide or constraint for a system that takes autonomous decisions[3].

Following the "principlism" approach, The ETHAI model was at first based on a set of ethics requirements.

Usually, when working in the field of biomedical ethics, the most common approach is to base any ethical evaluation on the famous four principles of medical bioethics, formulated in 1979 by Beauchamp and Childress. However, there is much debate on whether these principles still represent everything that counts and should be considered when deriving an ethical model for biomedical research. Huxtable R. (2001) for instance argues that the principles can only mark the beginning of the moral work, without managing to be conclusive, while Takala T. (2001) stresses the fact that the principles are too widely interpretable and so fail to provide a true "hands-on" foundation for global bioethics. Shea M. (2020) dedicated an interesting paper on the debate about the persisting (or not) validity of the four principles in today's bioethical landscape.

In order to validate it, the model based on the four principles was initially examined together with medical partners. The whole ETHAI model will afterwards be checked with technical partners to define, at the end of the project, a set of ethical requirements grounded in the real medical practice and with a solid technical feasibility, linked - following the approach of Morley et al (2020) - also to the different project and system development phases.

However, in parallel to the first application of the model with the medical partners, and partially as a result of the aforementioned activity, a reflection come up on how to preserve the doctor's and patient's autonomy (and if it is in fact really put at risk by the introduction of AI) and on how to incorporate into the requirements (and so into the system, by design) the point of view, the needs and the expectations of the patients.

This, along with all the discussions on the representativeness of the common approach based on the four principles, led us to critically reconsider them at the heart of the ETHAI model, shifting the attention towards an approach that puts the concept of care at its basis, on the same level as the other four principles, or even in a dominant position.

Moreover, following Groot B.C. et al (2019) and McCarty J. (2003) it is our conviction that a purely principlist position cuts out some fundamental aspects that should be included in any model that deals with the relationship between AI and healthcare, especially in the light of the possible dehumanization effects that come with the introduction of an "algorithmic" approach to medical research and practice.

Indeed, even if a purely principlist approach can satisfy all the principles (beneficence, non-maleficence, autonomy and justice) as well as the AI HLEG guidelines for trustworthy AI, it may still fail to handle some basic human needs that should come to the fore when dealing with health connected problems.

Interesting in this the position of Shea M. (2000), who argues that without an account of human well-being the principlism theory is not sufficient to orient the behavior and moral choices of healthcare professionals. Also interesting the position of Walker, T. (2009). who argues that principlism fails to provide a framework to help healthcare professionals to decide what to do in moral complex situations. Finally, it should be considered the position of Page, K. (2012) who highlights how people do not in fact actually use the principles in the decision-making process.

For this reason, a tailored set of requirements were developed, incorporating the philosophical concept of care (in this following the approach of Tronto, J. C. (2013)) and of Gilligan, C. (1982), and declining them in four nuances (caring about, caring for, care giving, care receiving) and adding the principle of Care at the basis of the model, together with the four other principles.

We might also conclude that such values and requirements should be incorporated into any system dealing in general with human beings and

[3] There are many different kinds of EDMs and theoretical approaches to them. A good comprehensive scoping review of the scientific literature about this can be found in Cottone R, Claus R.(2000) and in Melanie K.et al (2021)

their well-being, in healthcare or elsewhere. This should even constitute the keystone of any ethical decision model that in any way deals with the consequences of the adoption of a form of AI into a process that formerly was typically human.

5. Validation of the ETHAI model

Can we inscribe AI into a true "care" paradigm?

To answer this question, the research team is lucky enough to have the opportunity of developing, testing and validating the model in the context of MES-CoBraD in a real medical setting, allowing to i) work with developers and to embed the deriving requirements into the system, ii) apply the model to true medical practice occurring during the research, and iii) verify if the resulting decisions seem to be ethically sound.

With this purpose, the validity of the model itself will be verified all along the project, testing two different versions of the model on the same use cases, one based on the sole four principles of medical bioethics and another one with the addition of the care principles at its basis. In this way, it is expected to be able to compare the results to check if the two different models lead to different decisions and if the adopted approach leads to the desired outcome.

6. Conclusion

This paper presented the first results of a still ongoing research, where a humanistic ethical decision model for healthcare is under development. It highlights the necessity of a more humanistic approach to the ethics of AI - especially when it comes to AI applied to healthcare - and it is aimed to ascertain what model(s) of care is possible to successfully integrate AI mechanisms.

In order to support healthcare professionals and researchers in their daily decision-making process, and to guide system developers to conceive a system that delivers ethically sound outputs, a project tailored ethical decision model (EDM) called ETHAI was defined by the consortium partner CyberEthics Lab. in the context of the MES-CoBraD project, along with a set of ethics requirements. The ETHAI follows the ethics of care positions, adopting at its basis specific human values and characteristics, like empathy, caring, being in relation with other human beings. It will be applied by the project in the next months to carry out the assessment of the MES-CoBraD's AI-based technology.

7. Acknowledgements

This work is part of the MES-CoBraD project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 965422

8. References

[1] Brannelly, T. (ed.), Ward, L. (ed.), and Ward, N. (ed.), Ethics of Care: Critical advances in international perspective, online ed., Policy Press Scholarship Online, Bristol, 2016.

[2] Beauchamp, T.L. and Childress, J.F., Principles of Biomedical Ethics. 8th Edition, Oxford University Press, New York, 2013.

[3] T.L. Beauchamp T. J. Childress, Principles of Biomedical Ethics: Marking Its Fortieth Anniversary, The American Journal of Bioethics, 19:11, (2019) 9-12. doi: 10.1080/15265161.2019.1665402I

[4] Cottone, R.R., Claus, R.E., Ethical decision-making models: a review of the literature. J Couns Dev, 78:3 (2000) 275-83. doi: 10.1002/j.1556-6676.2000.tb01908.

[5] EPRS | European Parliamentary Research Service Scientific Foresight Unit Artificial intelligence in healthcare: Applications, risks, and ethical and societal impacts. (STOA) PE 729.512 – June 2002.

[6] Floridi, L., Matthias, H., Taddeo, M., Silva, J., Mökander, J., Wen, Y., capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act (March 23, 2022).

[7] Gilligan, C., In a different voice: Psychological theory and women's development. Harvard University Press, 1982.

[8] Groot, B. C., Vink, M., Haveman, A., Huberts, M., Schout, G. & Abma, T., Ethics of care in participatory health research: mutual responsibility in collaboration with co-researchers, Educational Action Research, 27:2 (2019) 286-302, doi. 10.1080/09650792.2018.1450771.

[9] Huxtable, R., For and against the four principles of biomedical ethics, Clinical

Ethics, 8:2-3 (2018) 39-43. doi:10.1177/1477750913486245.

[10] Johnson, K. K., Weeks, S. N., Gimpel Peacock, G., Domenech Rodríguez, M., Ethical decision-making models: a taxonomy of models and review of issues, Ethics & Behavior (2021). doi: 10.1080/10508422.2021.1913593.

[11] Karimian, G., Petelos, E. & Evers, S. M., The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review, AI Ethics 2 (2022) 539–551. https://doi.org/10.1007/s43681-021-00131-7.

[12] Lawrence, D. J., The Four Principles of Biomedical Ethics: A Foundation for Current Bioethical Debate, Journal of Chiropractic Humanities, 14 (2007) 34-40. https://doi.org/10.1016/S1556-3499(13)60161-8.

[13] G. Maio, Fundamentals of an Ethics of Care, In: F. Krause, J. Boldt, (ed) Care in Healthcare. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-61291-1_4.

[14] Matthew, S., Forty Years of the Four Principles: Enduring Themes from Beauchamp and Childress, The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 45:4-5 (2020) 387–395. https://doi.org/10.1093/jmp/jhaa020.

[15] McCarthy, J., Principalism or narrative ethics: must we choose between them? Medical Humanities 29 (2003) 65-71. doi: 10.1136/mh.29.2.65.

[16] Morley, J., Floridi, L., Kinsey, L., et al., From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices, Sci Eng Ethics 26 (2020) 2141–2168. https://doi.org/10.1007/s11948-019-00165-5.

[17] Morley, J., Machado, C., Burr, C., Cowls, J., Taddeo, M., Floridi, L., The Debate on the Ethics of AI in Health Care: A Reconstruction and Critical Review, 2019. http://dx.doi.org/10.2139/ssrn.3486518.

[18] Murphy, K., Di Ruggiero, E., Upshur, R., et al., Artificial intelligence for good health: a scoping review of the ethics literature, BMC Med Ethics 22:14 (2021). https://doi.org/10.1186/s12910-021-00577-8

[19] Oviedo Bioethics Convention, Convention for the Protection of Human Rights and Dignity of the Human Being, Oviedo, 4 April 1997. https://www.coe.int/en/web/bioethics/oviedo-convention

[20] U.S. Department of Health and Human Services, National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, The Belmont report: Ethical principles and guidelines for the protection of human subjects of research. 1979.

[21] Ostherr, K., Artificial Intelligence and Medical Humanities, J Med Humanit. 43:2 (2022) 211-232. doi: 10.1007/s10912-020-09636-4.

[22] Page, K., The four principles: can they be measured and do they predict ethical decision making? BMC Med Ethics, 13:10 (2012). doi: 10.1186/1472-6939-13-10.

[23] Ruckenstein, M., and Dow Schüll, N., The Datafication of Health, Annual Review of Anthropology 46:1 (2017) 261-278. https://doi.org/10.1146/annurev-anthro-102116-041244.

[24] Shea, M., Principlism's Balancing Act: Why the Principles of Biomedical Ethics Need a Theory of the Good, The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine, 45:4-5 (2020) 441–470. https://doi.org/10.1093/jmp/jhaa014.

[25] Takala, T., What Is Wrong with Global Bioethics? On the Limitations of the Four Principles Approach, Cambridge Quarterly of Healthcare Ethics, 10:1 (2021) 72-77. doi:10.1017/S0963180101001098.

[26] J. C. Tronto Caring Democracy: Markets, Equality, and Justice, New York University Press, New York, NY, 2013.

[27] Varkey, B., Principles of Clinical Ethics and Their Application to Practice. Med Princ Pract. 30:1 (2021) 17-28. doi: 10.1159/000509119.

[28] Veatch, R. M., Reconciling Lists of Principles in Bioethics, J Med Philos 29;45:4-5 (2020) 540-559. doi: 10.1093/jmp/jhaa017.

[29] Walker, T., What principlism misses. Journal of medical ethics. 35 (2009) 229-31. http://dx.doi.org/10.1136/jme.2008.027227.